DITA, XLIFF and Translation: Truths, Myths and Misconceptions

Andrzej Zydroń MBCS CITP
CTO
XML-INTL
Leaders in Translation Technology
September 2008
azydron@xml-intl.com
DITA is cheaper to translate: Truths

- Separation of content and format
- Component based architecture
- You only translate changed topics
Translation is the main cost of localization: Myths

Source: Professor Reinhard Schäler LRC - ASLIB 2002
DITA Granularity: a double edged sword

- DITA without a CMS – you must be NUTS!
- You DO NOT NEED a native XML CMS
- Links, links, links, links…..
- Translate topics as soon as they are available
- Increased project management costs
- Necessitate web services based exchange
- Need to establish long term relationship with Localization Service Providers
DITA Translation Pitfalls

- DITA comes ready packed with some very dangerous options
- Translatable acronyms
- Beware the CONREF for it may TRIPLE your translation costs
- Specialize if you dare
- DITA Translation TC Best Practices
DITA + XLIFF
Only part of the picture

SRX
W3C ITS
Unicode TR29
TMX
GMX
XLIFF

Author Memory
Translation Memory

xml:tm
XML Vocabulary, e.g. DITA
XML1.0
Unicode 5.0
OAXAL: OASIS Reference Architecture TC

- SRX
- Unicode TR 29
- GMX-V
- W3C ITS
- xml:tm
- DITA/XML
- XLIFF
- TMX
OAXAL TC

http://www.oasis-open.org/committees/tc_home.php?wg_abbrev=oaxal
Collaborative review and approval
Pre-translate content and alert translators
Vendor localizes XML content with DocZone translation tool
Write/edit XML content
Check content into DocZone server
Publish to all output formats for all markets
Create graphics
Store in DocZone
Link with XML content
DocZone.com example
OAXAL: Why is DITA + XLIFF not enough

- **Process Automation**
  - 50% Translation costs process management

- **Matching**
  - Strange commercial model for translation companies
  - Automation, automation, automation
DITA + OAXAL putting it together:

- Unicode TR 29
- SRX
- W3C ITS
- DITA/XML
- DITA/XML + xml:tm
Example of the use of `tm` namespace in an XML document:

```xml
<document xmlns:tm="urn:xml-intl-tm" te="9">
  <tm:tm>
    <section>
      <para>
        <tm:te id="e1">
          <tm:tu id="u1.1">
            Namespace is very flexible.
          </tm:tu>
          <tm:tu id="u1.2">
            It is very easy to use.
          </tm:tu>
        </tm:te>
      </para>
    </section>
  </tm:tm>
</document>
```
xml:tm namespace

Source document view

doc

<table>
<thead>
<tr>
<th>title</th>
<th>text</th>
</tr>
</thead>
<tbody>
<tr>
<td>section</td>
<td></td>
</tr>
<tr>
<td>para</td>
<td>text</td>
</tr>
<tr>
<td>para</td>
<td>text</td>
</tr>
<tr>
<td>para</td>
<td>text</td>
</tr>
</tbody>
</table>

Source document tm namespace view

tm

<table>
<thead>
<tr>
<th>te</th>
<th>tu</th>
<th>text</th>
</tr>
</thead>
<tbody>
<tr>
<td>section</td>
<td></td>
<td></td>
</tr>
<tr>
<td>para</td>
<td>text</td>
<td></td>
</tr>
<tr>
<td>para</td>
<td>text</td>
<td></td>
</tr>
<tr>
<td>para</td>
<td>text</td>
<td></td>
</tr>
<tr>
<td>para</td>
<td>text</td>
<td></td>
</tr>
<tr>
<td>para</td>
<td>text</td>
<td></td>
</tr>
<tr>
<td>para</td>
<td>text</td>
<td></td>
</tr>
</tbody>
</table>
xml:tm namespace

- Author memory
  - Maintain memory of source text
  - Authoring statistics
  - Authoring tool input
- Translation memory
  - Automatic alignment
  - Maintain exact link of source and target text
  - Reduce translation costs
xml:tm differencing

Original Source Document

tu id="1"
tu id="2"
tu id="3"
tu id="4"
tu id="5"
tu id="6"

DOM Differencing

delte
modified

Updated Source Document

tu id="1"
tu id="2"
tu id="3"
tu id="4"
tu id="6"
tu id="7"
tu id="8"
xml:tm author memory

- Namespace aware DOM differencing
- Identify changes from the previous version
- Unique text unit identifiers are maintained
- Modification history
- Text units can be loaded into a database
- Authoring environment integration
xml:tm author memory

- Namespace aware DOM differencing
- Identify changes from the previous version
- Unique text unit identifiers are maintained
- Modification history
- Text units can be loaded into a database
- Authoring environment integration
**Alarm**

[Prolog]

[Body]

The car is equipped with a fully installed Clifford alarm system that meets category 1 Thatchem requirements. The alarm can be operated with the Spyker design key-fob. With this key fob, the owner can arm/disarm the alarm and open the doors and trunk.

---

Figure:

- Key = Arm/disarm
- L = Open LH door
- R = Open RH door
XLIFF + xml:tm:
DITA/OAXAL to XLIFF

Original Source Document

- tu id="1"
- tu id="2"
- tu id="3"
- tu id="4"
- tu id="5"
- tu id="6"

XLIFF File

- Trans-unit id="1"
- Trans-unit id="2"
- Trans-unit id="3"
- Trans-unit id="4"
- Trans-unit id="5"
- Trans-unit id="6"

Translated Target Document

- tu id="1"
- tu id="2"
- tu id="3"
- tu id="4"
- tu id="5"
- tu id="6"
xml:tm exact matching

Updated Source Document

- tu id="1"
- tu id="2"
- tu id="3"
- tu id="4"
- tu id="7"
- tu id="6"
- tu id="8"

Exact Matching

- tu id="1"
- tu id="3"
- tu id="4"
- tu id="7"
- tu id="6"
- tu id="8"

Matched Target Document

- Exact match
- Exact match
- Exact match
- Exact match
- Exact match
- Exact match

deleted
modified
new
xml:tm matching

Updated Source Document

- tu id="1"
- tu id="2"
- tu id="3"
- tu id="4"
- tu id="7"
- tu id="6"
- tu id="8"
- tu id="9"

Matched Target Document

- tu id="1"
- tu id="2"
- tu id="3"
- tu id="4"
- tu id="7"
- tu id="6"
- tu id="8"
- tu id="9"

- non trans
- modified
- new:same

- requires no translation
- requires translation
- requires proofing
- fuzzy match origid="5"
- doc leveraged match
- DB leveraged match

Exact match
- non translatable
- Exact match
- Exact match
- fuzzy match origid="5"
- doc leveraged match
- DB leveraged match
xml:tm translated document

Translated document view:
- doc
  - title
    - tekst
- section
  - para
    - tekst
  - para
    - tekst
  - para
    - tekst

Translated document tm namespace view:
- tm
  - te
  - tu
  - tekst
  - zdanie
  - tu
  - zdanie
  - tu
  - zdanie
  - tu
  - zdanie
  - tu
  - zdanie
  - tu
  - zdanie
  - tu
  - zdanie
  - tu
  - zdanie
  - tu
  - zdanie
Translation without OAXAL:

source text -> source text -> extract -> extracted text -> tm process

source text -> target text

QA -> translated text

prepared text -> merge -> target text

translated text -> target text
OAXAL in action

Automated Workflow:
- xml:tm source text → extract
- extracted text → tm process
- XLIFF file
- exact matching
- leveraged matching

Internet:
- QA → web browser → translate

Automated Workflow:
- xml:tm target text → merge

exact matching
<table>
<thead>
<tr>
<th>ID</th>
<th>English, US</th>
<th>Japanese</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Changing the oil in your car.</td>
<td>オイル交換</td>
<td>M</td>
</tr>
<tr>
<td>2</td>
<td>Once every 6000 kilometers or three months, change the oil in your car.</td>
<td>走行6000キロごと、または3ヶ月ごとにオイルを交換しましょう。</td>
<td>M</td>
</tr>
<tr>
<td>3</td>
<td>This will help maintain the engine in good condition.</td>
<td>This will help maintain the engine in good condition.</td>
<td>RT</td>
</tr>
<tr>
<td>4</td>
<td>To change the oil:</td>
<td>オイル交換手順を示します。</td>
<td>M</td>
</tr>
<tr>
<td>5</td>
<td>Remove the old oil filter.</td>
<td>オイル・フィルターを取り外します。</td>
<td>M</td>
</tr>
<tr>
<td>6</td>
<td>Drain the oil.</td>
<td>オイルを抜き取ります。</td>
<td>M</td>
</tr>
<tr>
<td>7</td>
<td>Install a new oil filter and gasket.</td>
<td>新しいオイル・フィルターとガスケットを取り付けます。</td>
<td>M</td>
</tr>
<tr>
<td>8</td>
<td>Add new oil to the engine.</td>
<td>エンジンに新しいオイルを充填します。</td>
<td>M</td>
</tr>
<tr>
<td>9</td>
<td>Check the air filter and replace or clean it.</td>
<td>エア・フィルターを点検し、交換または洗浄します。</td>
<td>M</td>
</tr>
<tr>
<td>10</td>
<td>Top up the windshield washer fluid.</td>
<td>ウィンドウ洗浄液を補充します。</td>
<td>M</td>
</tr>
</tbody>
</table>
The garage is a good place to spray paint.

Move the car out of the garage to avoid getting paint on it.

Place newspaper, cardboard, or a drop-cloth on the garage floor.

Place the object to be painted within the covered area.

Follow the directions on the paint can to paint the object.

Let the paint dry thoroughly before you move the object.

Remove all articles placed on the floor and dispose of safely.
Spray painting

The garage is a good place to spray paint.

Move the car out of the garage to avoid getting paint on it.

Place newspaper, cardboard, or a drop-cloth on the garage floor.

Place the object to be painted within the covered area.
<xml version="1.0" encoding="UTF-8"/>

<!DOCTYPE html>

<task id="spraypaint" xml:lang="en-us" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
     xsi:noNamespaceSchemaLocation="../schema/task.xsd">
  <title>Spray painting</title>
  <taskbody>
    <p>The garage is a good place to spray paint.</p>
  </taskbody>
  <steps>
    <step><cmd>Move the car out of the garage to avoid getting paint on it.</cmd></step>
    <step><cmd>Place newspaper, cardboard, or a drop-cloth on the garage floor.</cmd></step>
    <step><cmd>Place the object to be painted within the covered area.</cmd></step>
    <step><cmd>Follow the directions on the paint can to paint the object.</cmd></step>
    <step><cmd>Let the paint dry thoroughly before you move the object.</cmd></step>
    <cmd>Remove all articles placed on the floor and dispose of safely.</cmd>
  </steps>
  <related-links>
    <link format="xml" href="../concepts/paint.xml" type="concept">
      <linktext>Paint</linktext>
    </link>
  </related-links>
</task>
- <trans-unit id="t1" translate="no" ts="title" reformat="yes" xml:space="default" datatype="normal">
  <source>Spray painting</source>
  <target ts="matched">スプレー式のペンキ</target>
</trans-unit>
- <trans-unit id="t2" translate="no" ts="p" reformat="yes" xml:space="default" datatype="normal">
  <source>The garage is a good place to spray paint.</source>
  <target ts="matched">ガレージはスプレー式のペンキに適しています。</target>
</trans-unit>
- <trans-unit id="t3" translate="no" ts="cmd" reformat="yes" xml:space="default" datatype="normal">
  <source>Move the car out of the garage to avoid getting paint on it.</source>
  <target ts="matched">ペンキが付着しないように車をガレージから出します。</target>
</trans-unit>
- <trans-unit id="t4" translate="no" ts="cmd" reformat="yes" xml:space="default" datatype="normal">
  <source>Place newspaper, cardboard, or a drop-cloth on the garage floor.</source>
  <target ts="matched">ガレージの床に、新聞、ダンボールまたは汚れ防止の布を敷きます。</target>
</trans-unit>
- <trans-unit id="t5" translate="yes" ts="cmd" reformat="yes" xml:space="default" datatype="normal">
  <source>Place the object to be painted within the covered area.</source>
  <target ts="translate">Place the object to be painted within the covered area.</target>
  <alt-trans ts="indoc-fuzzy">
    <source tsp="fuzzy">Place the object to be painted on the covered area.</source>
    <target tsp="fuzzy">敷き詰めた汚れ防止の上に、ペイントする物を置きます。</target>
  </alt-trans>
</trans-unit>
Contact Details

- Postal address:
  - PO Box 2167
  - Gerrards Cross
  - Bucks SL9 8XF
  - United Kingdom

- Phone: +44 1753 480 467
- Fax: +44 1753 480 465
- Andrzej Zydroń – azydron@xml-intl.com